There now follows, by way of example only, a description of specific embodiments of the present invention. The description is given with reference to the accompanying drawings in which:

Figure 1 shows the hardware used in providing a first embodiment of the present invention;

Figures 2A and 2B show the top-level design of a text-to-speech conversion program which controls the operation of the hardware shown in Figure 1;

Figures 3A & 3B show the text analysis process of Figure 2A in more detail;

Figure 4 is a diagram showing part of a syntactic classification of words; and

Figure 5 is a flow chart illustrating the prosodic structure assignment process of Figure 2B.

Figure 1 shows a hardware configuration of a personal computer operable to provide a first embodiment of the present invention. The computer has a central processing unit 10 which is connected by data lines to a Random Access Memory (RAM) 12, a hard disc 14, a CD-ROM drive 16, input/output peripherals 18,20,22 and two interface cards 24,28. The input/output peripherals include a visual display unit 18, a keyboard 20 and a mouse 22. The interface cards comprise a sound card 24 which connects the computer to a loudspeaker 26 and a network card 28 which connects the computer to the Internet 30.

The computer is controlled by conventional operating system software which is transferred from the hard disc 14 to the RAM 12 when the computer is switched on. A CD-ROM 32 carries:

a) software which the computer can execute to provide the user with a text-to-speech facility; and

b) five databases used in the text-to-speech conversion process.

To use the software, the user loads the CD-ROM 32 into the CD-ROM drive 16 and then, using the keyboard 20 and the mouse 22, causes the computer to copy the software and databases from the CD-ROM 32 to the hard disc 14. The user can then select a text-representing file (such as an e-mail loaded into the computer from the Internet 30) and run the text-to-speech program to cause the computer to produce a spoken version of the e-mail via the loudspeaker 26. On running the text-to-speech program both the program itself and the databases are loaded into the RAM 12.

The text-to-speech program then controls the computer to carry out the functions illustrated in Figures 2A and 2B. As will be described in more detail below, the computer first carries out text analysis process 42 on the e-mail (shown as text 40) which the user has indicated he wishes to be converted to speech. The text analysis process 42 uses a lexicon 44 (the first of the five databases stored on the CD-ROM 32) to generate word grouping data 46, syntactic information 48 and phonetic transcription data 49 concerning the text-file 40. The output data 46,48,49 is stored in the RAM 12.

After completion of the text analysis program 42, the program controls the computer to carry out the prosodic structure prediction process 50. The process 50 operates on the syntactic data 48 and word grouping data 46 stored in RAM 12 to produce phrase boundary data 54. The phrase boundary data 54 is also stored in RAM 12. The prosodic structure prediction process 50 uses the prosodic structure corpus 52 (which is the second of the five databases stored on the CD-ROM 32). The process will be described in more detail (with reference to Figures 4 and 5) below.

Once the phrase boundary data 54 has been generated, the program controls the computer to carry out prosody prediction process (Figure 2B, 56) to generate performance data 58 which includes data on the pitch, amplitude and duration of phonemes to be used in generating the output speech 72. A description of the prosody prediction process 56 is given in Edgington M et al: 'Overview of current text-to-speech techniques part 2 – prosody and speech synthesis', BT Technology Journal, Volume 14, No. 1, pp 84-99 (January 1996). The disclosure of that paper (hereinafter referred to as part 2 of the BTTJ article) is hereby incorporated herein by reference.

Thereafter, the computer performs a speech sound generation process 62 to convert the phonetic transcription data 49 to a raw speech waveform 66. The process 62 involves the concatenation of segments of speech waveforms stored in a speech waveform database 64 (the speech waveform database is the third of the five databases stored on the CD-ROM 32). Suitable methods for carrying out the speech sound generation process 62 are disclosed in the applicant's European patent no. 0 712 529 and European patent application no. 95302474.9. Further details of such methods can be found in part 2 of the BTTJ article.

Thereafter, the computer carries out a prosody and speech combination process 70 to manipulate the raw speech waveform data 66 in accordance with the performance data 58 to produce speech data 72. Again, those skilled in the art will be able to write suitable software to carry out combination process 70. Part 2 of the BTTJ article describes the process 70 in more detail. The program then controls the computer to forward the speech data 72 to the sound card 24 where it is converted to an analogue electrical signal which is used to drive loudspeaker 26 to produce a spoken version of the text file 40.

The text analysis process 42 is illustrated in more detail in Figures 3A and 3B. The program first controls the computer to execute a segmentation and normalisation process (Figure 3A, 80). The normalisation aspect of the process 80 involves the expansion of numerals, abbreviations, and amounts of money into the form of words, thereby generating an expanded text file 88. For example, '£100' in the text file 40 is expanded to 'one hundred pounds' in the expanded text file 88. These operations are done with the aid of an abbreviations database 82, which is the fourth of the five databases stored on the CD-ROM 32. The segmentation aspect of the process 80 involves the addition of start-of-sentence, end-of-sentence, start-of-paragraph and